

PREDICTION LUNG CANCER BASED CRITICAL FACTORS USING MACHINE LEARNING

Sherko H. Murad^{a*}, Ardalan H. Awlla^b, Brzu T. Mohammed^c

^{a, c} Computer Science Department, Kurdistan Technical Institute, Sulaimani, Kurdistan Region, Iraq - sherko.murad@kti.edu.iq

^b Department of Computer Science, Cihan University Sulaimaniya, Sulaymaniya 46001, Kurdistan Region, Iraq

ardalan.husin@gmail.com

Received: 11 Jan., 2023/ Accepted: 14 May, 2023/ Published: 25 Sep., 2023

<https://doi.org/10.25271/sjuoz.2023.11.3.1105>

ABSTRACT:

Many people around the world have lung cancer. Lung cancer has a poor prognosis and a high mortality rate. Through image recognition and data analytics, computers can play a significant role in detecting various types of cancer disease. This paper provides an effective method to predict lung cancer in an early stage with high accuracy ratio. This research proposed data analytics to determine the accuracy ratio of lung cancer patients using supervised machine learning algorithms (Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF)). The dataset for this study was obtained from "Data World," which contains 1,000 diseases. Machine learning algorithms enable us to identify lung cancer risk factors, which aid in diagnosing lung cancer. This study shows that those algorithms can classify lung cancer patients, with Random Forest having the highest accuracy of 98.507%.

KEYWORDS: Lung cancer, machine learning, SVM, DT, and RF.

1. INTRODUCTION

Cancer can manifest itself in various parts of the human body. Even for a long time, the disease may go unnoticed. According to the WHO, cancer can be avoided if this is the case. Early, adequate recognition (Chaudhori et al., 2021; Khorshid et al., 2021; Fairouz et al., 2021; Zeebaree et al., 2021; Dahkaz et al., 2021; Chauhau et al., 2016). One of the cancer types that can ultimately cause death is lung cancer, which is also one of the more well-known cancer types. However, if detected early, it is predicted that 15% of lung cancer patients receiving therapy will live for more than five years after their diagnosis. (Junior et al., 2018). By looking at these parameters, a computer can help diagnose lung cancer. Lung cancer has the greatest mortality rate of all of these illnesses. This disease is predicted to kill approximately 1.7 million people annually (Abdulqade et al., 2020). Lung cancer has a dismal prognosis and is greatly influenced by the tumor's stage at diagnosis. The two clinically-treated lung cancer types are non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC) (Ibrahim et al., 2020; Singh et al., 2019; Somvansh et al., 2016). In actuality, it is a malignant tumor marked by the development of cellular tissue that is out of sequence. When cancer cells invade new tissues, the process is known as metastasis. Cancer tends to spread and is incurable if it goes too far; thus, it should be found as early as possible. Lung cancer only shows symptoms in its advanced stages, making it challenging to identify and practically impossible to treat at this stage. Images of the lungs are captured using imaging techniques such as computed tomography (CT), positron emission tomography (PET), magnetic resonance imaging (MRI), and X-ray. The most widely used imaging method is the CT image technique since it can provide a view without overlapping components. Doctors have a difficult time interpreting and recognizing cancer.

Machine learning is required for complex data categorization and decision-making (Faisal et al., 2018; Zeebaree et al., 2018).

Machine learning is classified into two types: supervised machine learning and unsupervised machine learning. Many systems have insufficient detection accuracy, and particular systems must be constructed to attain the highest level of precision. Machine learning and image-processing approaches were used to detect and classify lung cancer (Abdulqader et al., 2020). However, some characteristics of lung cancer patients, such as smoking behaviors, may contribute to the early diagnosis of the illness. (Zeebaree et al., 2019; Karhan et al., 2016; Gunaydin et al., 2019).

In this study, a supervised machine learning algorithm with a high accuracy ratio, such as Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF), was used to predict lung cancer risk factors. With the help of machine learning, we can detect which factor has the most effect on lung cancer. For example (smoking, dry cough, passive smoking, long-term lung disease, dust mite allergy, genetic possibility, and being overweight). The dataset was given by (Data World). The ongoing research contributes to developing machine learning algorithms that detect lung cancer more accurately.

2. LITERATURE REVIEW

Machine learning uses several types of algorithms. While each of these algorithms approaches data differently, this section will present a few freshly proposed machine-learning methods in the field of lung cancer.

(Abdullah et al., 2020). this paper examined the accuracy ratio of three classifiers, namely Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Convolutional Neural Network (CNN), to classify lung cancer in its early stages. Results show that SVM gives the best result with 95.56%, followed by CNN with 92.11% and KNN with 88.40%.

(Karahhan et al., 2016). Different outcomes are provided for each classifier on the lung cancer dataset used in this paper. Different classifiers were applied, and matching accuracy rates of 99.3% were obtained, with the Support Vector Machine being the best.

* Corresponding author

This is an open access under a CC BY-NC-SA 4.0 license (<https://creativecommons.org/licenses/by-nc-sa/4.0/>)

(Günaydin et al., 2019). Principal Component Analysis, K-NN, SVM, Nave Bayes, Decision Trees, and ANN were suggested as machine learning algorithms for identifying anomalies in lung cancer nodules. After that, the two methods were contrasted with and without preprocessing. According to the test results, Decision Tree produces the best results with 93,24% percent accuracy without image processing.

(Banerjee et al., 2020), proposed a tumor classification paradigm. When the precision is compared to the suggested model, it is clear that the accuracy has increased, but the recall has reduced. MATLAB R2017a was utilized for digital image analysis, and machine learning classification was performed in a Jupiter notebook. The accuracy for region-based features was 79% percent, SVM 86%, and ANN 92% percent.

(Roy et al., 2020). They use image processing, biological approaches, and data discovery to increase accuracy and determine precise significance for early lung cancer identification. The lungs' representation from a CT scan (Computer Tomography) The Region of Interest (ROI) is determined after preprocessing the scan pictures. The Random Forest method is used to separate the distinctive features. Using an SVM Classifier and the SURF (Speeded Up Robust Functionality) technique, characteristics like entropy, correlation, power, and variance were extracted from Saliency Enhanced pictures. Whether an image is hazardous or safe depends on its classification (carcinomic). There were CT scan images in the dataset. The process was completed using the random forest and SVM classification techniques.

(Elnakib et al., 2020) recommended early lung node identification utilizing low-dose computed tomography (LDCT) images. The suggested gadget initially transforms unprocessed data to enhance comparing low-dose videos. Then, it is looked into how well several architectures, including Alex, VGG16, and VGG19 networks, perform compact profound learning. A genetic algorithm (GA) is taught to recognize the most crucial early-finding features in order to optimize the generated collection of data. Then, different classifiers are evaluated to identify lung nodules accurately. With VGG19 and SVM classification, the suggested method obtains the best detection precision, 97.5% percent.

(Faisal et al., 2018). advocate testing machine learning classifiers for lung cancer detection alongside traditional classifiers. The dataset was taken from the UCI registry and is being examined for the prediction of lung cancer using random forest and plurality voting-based ensembles. All of the investigated individual and ensemble classifiers were outperformed by the gradient-boosted Tree. According to performance evaluations, the gradient-boosted Tree outscored all competitors and ensemble classifiers, obtaining 90% precision.

(Reddy et al., 2019). used machine learning algorithms to present a successful model for detecting lung cancer stages. The model combines K-NN, Decision Trees, and Neural Network structures with the bagging ensemble approach to improve overall prediction accuracy. The proposed model's estimated outcomes are more accurate than individual algorithms. The bootstrap aggregating methodology improves the performance of the individual models, with accuracy scores of 97% (Decision Tree), 94%, and 96% (K-NN), respectively (Neural Networks). The integrated model has an accuracy score of 0.98%. The integrated model's precision has increased by 3.33% percent.

3. DATA COLLECTION AND EXCREMENTAL FRAMEWORK

3.1 Data collection

Lung cancer data was used in this study, and it is available on the Data World cloud-native SaaS platform. The dataset collected includes 1,000 patient records and 21 attributes that describe the

signs and symptoms of lung cancer and its health conditions. Low, moderate, and significant risk levels are represented by the three main categories in the dataset. The dataset is examined to see how each feature affects estimating the level of danger.

3.2 Data preprocessing

Data cleaning, selection, and normalization are the stages of processing the dataset. During the data cleaning stage, a reliable data format is created to check missing data, identify duplicate data, and clean up insufficient data within the second step. To identify the most significant lung cancer features from a medical perspective, several medical surveys are issued to many doctors and experts in the field of lung cancer. Following data collection and factor extraction, medical surveys are developed based on the extracted 21 factors, as shown in Table 1. and distributed it among several lung cancer experts, whose expertise determined the risk level of each of the 21 factors ranging from 1 to 10 points, with the most effective receiving 10 points and the weakest receiving 1. See Table 1 for the detailed survey of the collected attributes. Finally, using the knowledge gained from the first and second steps, a Google Collab program is used to forecast the lung cancer risk based on the results of these two steps.

Table 1: Description of attributes used for predicting the lung cancer.

No	Features	Description	Importance Level (0-10)
1	Smoking	Does the patient smoke?	7
2	Dry Cough	Does the patient have a dry cough?	5
3	Passive smoker	Has the patient ever smoked?	8
4	Long-term lung disease	Does the patient have chronic lung disease?	3
5	Genetic possibility	Does the patient have genetic lung cancer?	5
6	Overweig ht	does the patient is	0
7	Air pollution	Is the patient's environment have pollution?	6
8	Alcohol use	How long has the patient been drinking alcohol?	5
9	Occupatio nal risk	Is the patient's work environment dangerous?	6
10	Balanced diet	is the patient eating healthily a balanced diet	6
11	Chest ache	Does the patient have chest pain?	3
12	Hemoptys is	Does the patient have bloody cough?	4
13	Tiredness	How does lung cancer cause fatigue?	4
14	Weight Loss	Is the patient losing weight	2
15	Asthmatic	Does the patient have shortness of breath?	4
16	Wheezing	does the patient have whistling noise while breathing?	5
17	Dysphagia	Does the patient have difficulty in	4

		swallowing?	
18	Digital clubbing	Finger clubbing can be a sign of lung cancer	4
19	Snoring	Does the patient have snoring while you sleep	5
20	Age	Patient's age	7
21	Gender	Patient's gender	0
22	Measure	Is there lung cancer or not?	

The paper suggests a model to predict and classify the lung cancer classes. The proposed model starts with data pre-processing, feature selection, classification, and evaluation. Figure 1. shows the block diagram for the proposed work.

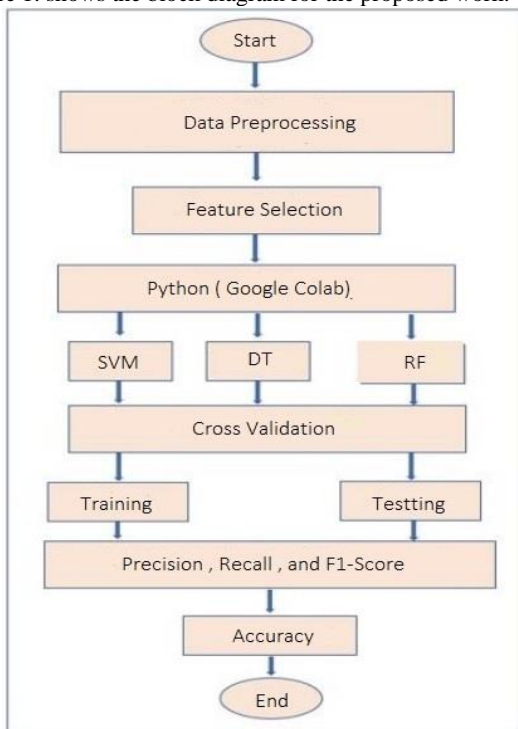


Figure 1: illustrates the block diagram for the proposed work.

3.3 Methodology

Medical prediction has recently benefited from using machine learning (ML) techniques. Numerous ML algorithms can be used for a wide range of applications. Numerous studies have demonstrated how ML algorithms have enhanced clinical support and decision-making based on patient data. One of the valuable and potent uses of ML forecast algorithms in the medical services industry is illness prediction assessment. In order to examine atypical datasets related to lung disease, a machine-learning approach was proposed in this study (Reddy et al.,2019). The death rate in the area was examined for this article using the criteria listed in Table (1).

The Support Vector Machine (SVM) supervised classification algorithm is frequently utilized for linear classification and regression issues. SVM can therefore resolve both linear and nonlinear issues. The kernel function for SVM is selected based on the points of the variables in the hyperplane, and it offers a singular and ideal solution. In the equation, $W.X. + b = 0$, where w is a weight vector, x is the characteristic value, and b is a scalar commonly referred to as bias, $W.X. + b = 0$ (Muhamed et al., 2021)

The Decision Tree (D.T.) technique is a form of supervised learning that may be applied to classification and regression

problems; however, the classification of problems is where it is most frequently utilized. It is a D.T. classifier, and the structure of this method can be broken down into three parts: internal nodes, which represent dataset features; branches, which exhibit rules; and leaves, which indicate the conclusion for each leaf in the classification.

The supervised learning method, Random Forest (RF), is used for classification, regression, and other tasks. It functions by building many decision trees during training and predicting the class represented by each tree's mean. One kind of tree structure is a decision tree (a binary or a non-binary tree). Each branch indicates the feature attribute's outcome over a range of values, and each leaf node contains a category, so each non-leaf node represents a feature test.

3.4 Data Classification

Supervised classification has been proposed as an efficient automated method for detecting lung cancer. Supervised learning often involves two advanced processes. In the first step, referred to as the learning step, the supervised classification model trains the training dataset during the learning phase to generate classification rules. The model is tested using a new dataset to determine its classification accuracy in the second step. The supervised classification's effectiveness is then validated by comparing the labelled samples to the new test data. If the proposed model's security is robust, it can classify new unlabelled datasets. Figure 2. is an illustration of the supervised classification model. Finally, classification model techniques, such as rule-based algorithms, decision trees, neural networks, and Bayesian techniques, can be used for classification.

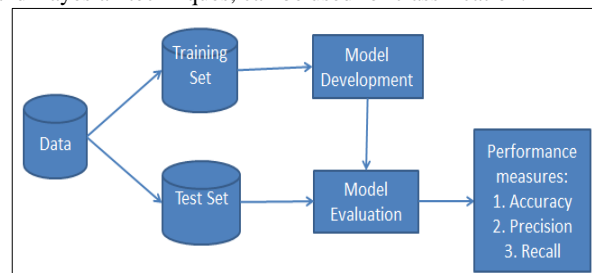


Figure 2: Illustration of supervised classification.

This study will clarify the outcomes of the proposed classification algorithms in this section. The Google Collab service, which makes a Tesla T4 GPU and 12 GB of RAM available to researchers worldwide, was used to implement the practical portion of this study in Python. Our model was tested using the provided web-based dataset.

3.5 Evaluation metrics

3.5.1 Confusion Matrix

The Confusion Matrix is a visual assessment method for deep learning. A Confusion Matrix's Figure 3. columns represent the prediction class results, while the rows represent the real class results. This matrix contains all the raw data about a classification model's assumptions on a given data set. To determine the accuracy of a model. It is a square matrix with the rows representing the actual class of the instances and the columns representing their expected class. When dealing with a binary, the confusion matrix is a 2 x 2 matrix that reports the number of true positives (TP), true negatives (T N), false positives (FP), and false negatives (F N).

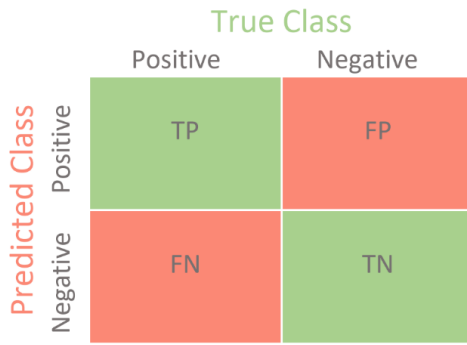


Fig 3: Confusion Matrix

The Confusion Matrix is a visual assessment method for deep learning. A Confusion Matrix's columns represent the prediction class results, while the rows represent the real class results. This matrix contains all the raw data about a classification model's assumptions on a given data set. To determine the accuracy of a model. It is a square matrix with the rows representing the actual class of the instances and the columns representing their expected class. When dealing with a binary, the confusion matrix is a 2 x 2 matrix that reports the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

To determine the accuracy level of our model, the precision of the parameter (1), recall (2), also known as responsiveness, and F1 score (3) have been used to compute the accuracy. The sum of the true positives and false negatives is divided by the number of true positives; this refers to the study's ability to recognize people with the condition who are ill correctly. In medical conditions, diseases are frequently categorized as a positive category. Neglecting this positive category has significant effects, including misdiagnosis, which could delay patient treatment. The diagnosis of medical lung cancer thus necessitates a high level of sensitivity or recall; precision (PPV) is required to determine how many of the expected positive cases are positive. When we want to recognize how many of the expected confirmed samples are positive, precision (PPV) is necessary. Accuracy is attained by dividing the total number of true positives and false positives by the sum of true positives and false positives. Accuracy in the field of lung cancer is required. All models F1 scores (3) are calculated using their corresponding sensitivity and precision. By using the following equations, accuracy, sensitivity, precision, and F1 score are determined:

$$Accuracy = \frac{TN + TP}{TN + FP + TP + FN}$$

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{FN + TP} \quad (2)$$

$$F1\ Score = 2 * \frac{Recall * Precision}{Recall + Precision} \quad (3)$$

4. RESULT AND DISCUSSION

In this study, we applied three classification algorithms to the lung cancer dataset: random forest, decision tree, and support vector machines (SVM). Then, the accuracy of each of those techniques was calculated using the metrics mentioned above. The accuracy percentages for each of those methods are displayed in Table (2).

Table 2: Performance prediction accuracy

Classifiers	Accuracy	Class	Precision	Recall	F1-Score
SVM	97.525%	Low	100%	98.795%	0.993938
		Medium	92.727%	100%	0.962263
		High	98.413%	93.939%	0.96124
Decision Tree	97.98%	Low	100%	98.795%	0.993938
		Medium	94.545%	100%	0.97196
		High	98.361%	95.238%	0.967743
Random Forest	98.507%	Low	100%	96.471%	0.982038
		Medium	98.148%	100%	0.990653
		High	96.923%	100%	0.984375

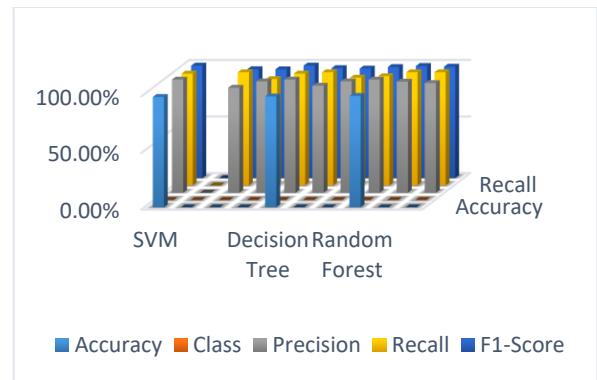


Figure 4: Proposed algorithms for performance prediction accuracy

The experimental results in Table 2 and Figure 4. shows that the Random Forest model has a better accuracy rate than the other models, achieving 98.507% percent. The results show that the Random Forest model has performed very well, with 100% recall for low classes, 98.148% for medium classes, and 96.923% for high classes. As for the recall, there is a 100% recall for both medium and high classes and 96.471.1% for the low class, which is the highest precision compared to the other models. Moreover, for an F1-score random forest, we have achieved a precise result, as shown in Table 2. Random Forest achieved 0.98% for the low class, 0.99% for the medium class, and 0.98% for the high class.

4.1 Analysis of Research Findings with Previous Study

After Compare the results of our study to the results of another study (Dahkaz et al., 2021), It is found that the algorithms in this study produce better results. By analysing patient data, an accurate cancer diagnosis can be made due to the use of influential factors to detect the disease. Table (3), shows the results.

Table 3. compares the research findings to previous studies.

Previous Study (Lung Cancer Dataset: UCI [5])	Accuracy	Our Study (Lung Cancer Dataset: data. World)	Accuracy
SVM	95.56	SVM	97.52
KNN	89.65	DT	97.98
CNN	92.11	RF	98.50

Table 4: Comparison of related work

Ref	Dataset	Classifier	Result
Karhan et al., 2016		SVM	99.3%
Günaydin et al.,2019	JSRT	k-NN, SVM, Nave Bayes, Decision Tree, ANN	The experimental results ANN 82,43% Decision Tree 93,24%
Banerjee et al., 2020	LIDC-LDRI	Random Forest, SVM and, ANN	Random Forest 70% SVM 80% ANN 96%
Roy et al.,2019	CT scan images	random forest algorithm and SVM classification	94.5%
Elnakib et al.,2020	LDCT images	SVM	96.25%
Faisal et al., 2018	UCI	(SVM), C4.5, Multi-Layer Perceptron, Decision tree, Naïve Bayes, and Neural Network	90%
Reddy et al., 2019	Data world source	DT, KNN, Neural Networks	Decision Tree 97% K-NN 94% Neural Networks
This work	UCI	SVM, DT, Random Forest	SVM 97.52% DT 97.98% Random Forest 98.50%

As shown in Table (4), the researchers used various methods, datasets, and feature selection/feature extraction methods. Compared to related work, we achieved a good result with the dataset and methods we used in this work.

4.2 The Results of random forest:

Because random forest produces accurate results, an RF algorithm generates random trees. To determine the most

practical features of lung cancer among all the features, as shown in Figure.5.

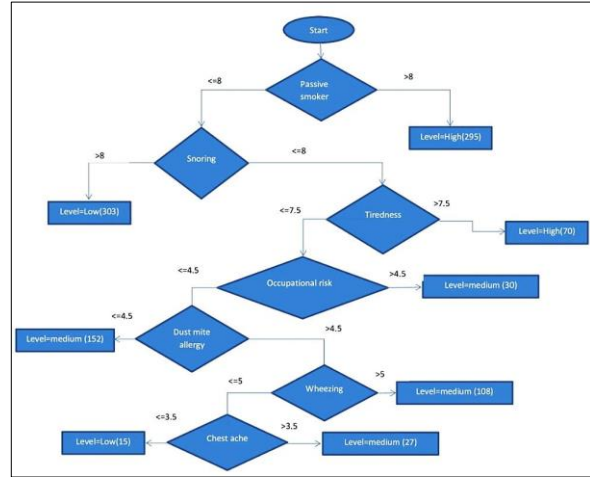


Figure.5. Generated random tree for the random forest algorithm based on the dataset

As illustrated in Figure 5. smoking is the leading cause of lung cancer, and people who smoke a pack of cigarettes or more daily for an extended period are more likely to develop lung cancer. Then two more influential factors cause lung cancer: tiredness and the workplace environment. Additionally, the type of work and the working environment are two other leading causes of lung disease, and the symptoms of these primary causes include shortness of breath and hoarseness in the throat during breathing; if the patient reaches these symptoms, it usually causes death.

5. CONCLUSION

The primary goal of this study is to raise public awareness of lung cancer risk factors so that people can regularly check their health when the risk is elevated. This study uses classification methods such as Decision Tree, SVM, and Random Forest. Random Forest achieved the highest accuracy among the three algorithms while training them on the lung cancer dataset. Therefore, we used the Radom Forest algorithm for analyzing the lung cancer dataset. We found that among the 21 causes and symptoms of lung cancer, long-term smoking is the leading cause of lung cancer, followed by work and work environment if a patient's workplace is filthy and dusty.

Acknowledgements: The authors would like to express their gratitude to the Computer Sciences department for supplying the laboratory for this study.

Conflict of interest: Sherko Murad. Ardalan Awlla and Brzu Tahir declares that he has no conflict of interest.

Authors contribution: A. conceived of the presented idea. A.B. developed the theory and performed the computations. C. Verified the analytical methods. A.B. supervised the findings of this work. All authors discussed the results and contributed to the final manuscript.

REFERENCES

Chaudhari, P., Agarwal, H., & Bhateja, V. (2021). Data augmentation for cancer classification in oncogenomics: an improved KNN based approach. *Evolutionary Intelligence*, 14, 489-498. Khorshid, S. Khorshid, S. F., & Abdulazeez, A. M. (2021). Breast cancer diagnosis based on k-nearest neighbors: a

- review. *PalArch's Journal of Archaeology of Egypt/Egyptology*, 18(4), 1927-1951.
- Fairouz Q. K., & Adnan M. A. (2021). Ultrasound Medical Images Classification Based on Deep Learning Algorithms. *Fusion: Practice and Applications, A Review*, vol. 3, no. 1, pp. 29–42, 2021, Doi: 10.5281/zenodo.4621289.
- Zeebaree, D. Q., Abdulazeez, A. M., Zebari, D. A., Haron, H., & Hamed, H. N. A. (2021). Multi-level fusion in ultrasound for cancer detection based on uniform LBP features. *Computers, Materials & Continua*, 66(3), 3363-3382.
- Dakhaz, M. A., Adnan, M. A. & Amira B.S. (2021). "Lung cancer prediction and classification based on correlation selection method using machine learning techniques", *Qubahan Academic Journal*, vol. 1.2, pp. 141-149.
- Chauhan, D., & Jaiswal, V. (2016). Development of computational tool for lung cancer prediction using data mining. *Int J Comput Appl Technol Res*, 5(17), 417-421.
- Abdullah, D. M., Abdulazeez, A. M., & Sallow, A. B. (2021). Lung cancer prediction and classification based on correlation selection method using machine learning techniques. *Qubahan Academic Journal*, 1(2), 141-149.
- Junior, J. R. F., Koenigkam-Santos, M., Cipriano, F. E. G., Fabro, A. T., & de Azevedo-Marques, P. M. (2018). Radiomics-based features for pattern recognition of lung cancer histopathology and metastases. *Computer methods and programs in biomedicine*, 159, 23-30.
- Ibrahim, I., & Abdulazeez, A. (2021). The role of machine learning algorithms for diagnosing diseases. *Journal of Applied Science and Technology Trends*, 2(01), 10-19.
- Singh, G. A. P., & Gupta, P. K. (2019). Performance analysis of various machine learning-based approaches for detection and classification of lung cancer in humans. *Neural Computing and Applications*, 31, 6863-6877.
- Somvanshi, M., Chavan, P., Tambade, S., & Shinde, S. V. (2016, August). A review of machine learning techniques using decision tree and support vector machine. In 2016 international conference on computing communication control and automation (ICCUBEA) (pp. 1-7). IEEE.
- Abdulqader, D. M., Abdulazeez, A. M., & Zeebaree, D. Q. (2020). Machine learning supervised algorithms of gene selection: A review. *Machine Learning*, 62(03), 233-244.
- Faisal, M. I., Bashir, S., Khan, Z. S., & Khan, F. H. (2018, December). An evaluation of machine learning classifiers and ensembles for early-stage prediction of lung cancer. In 2018 3rd international conference on emerging trends in engineering, sciences and technology (ICEEST) (pp. 1-4). IEEE.
- Zeebaree, D. Q., Haron, H., & Abdulazeez, A. M. (2018, October). Gene selection and classification of microarray data using convolutional neural network. In 2018 International Conference on Advanced Science and Engineering (ICOASE) (pp. 145-150). IEEE.
- Zeebaree, D. Q., Haron, H., Abdulazeez, A. M., & Zebari, D. A. (2019, April). Trainable model based on new uniform LBP feature to identify the risk of the breast cancer. In 2019 international conference on advanced science and engineering (ICOASE) (pp. 106-111). IEEE.
- Karhan, Z., & Tuğ, T. (2016). Lung Cancer Detection and Classification with Classification Algorithms. *IOSR Journal of Computer Engineering (IOSR-JCE)*, 18(6), 71-7.
- Günaydin, Ö., Günay, M., & Şengel, Ö. (2019, April). Comparison of lung cancer detection algorithms. In 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT) (pp. 1-4). IEEE.
- Banerjee, N., & Das, S. (2020, March). Prediction lung cancer– in machine learning perspective. In 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA) (pp. 1-5). IEEE.
- Roy, K., Chaudhury, S. S., Burman, M., Ganguly, A., Dutta, C., Banik, S., & Banik, R. (2019, March). A Comparative study of Lung Cancer detection using supervised neural network. In 2019 International Conference on Opto-Electronics and Applied Optics (Optronix) (pp. 1-5). IEEE.
- Elnakib, A., Amer, H. M., & Abou-Chadi, F. E. (2020). Early lung cancer detection using deep learning optimization.
- Muhammed, B. T., Awlla, A. H., Murad, S. H., & Ahmad, S. N. (2021). Prediction of CoVid-19 mortality in Iraq-Kurdistan by using Machine learning. *UHD Journal of Science and Technology*, 5(1), 66-70.
- Reddy, D., Kumar, E. N. H., Reddy, D., & Monika, P. (2019, July). Integrated Machine Learning Model for Prediction of Lung Cancer Stages from Textual data using Ensemble Method. In 2019 1st International Conference on Advances in Information Technology (ICAIT) (pp. 353-357). IEEE.